AN ANALYSIS OF WEIGHT GAIN AMONG CALL CENTER WORKERS
Statistical Analysis Plan
Xiao Liu, Gokcen Buyubas & Elizabeth Anderson

Introduction

Obesity related diseases are costly for businesses, incentivising interventions to reduce obesity. We examine factors related to weight gain in one specific case of employees working at one call center. First, we will examine multiple measures of weight gain, in addition to explanatory variables, such as the time of shift and exercise levels. To identify associations between factors, such as exercise and shift time, and weight gain, we use random forest regression and full information maximum likelihood estimation.

Analysis Population

Previous research identified female call center employees as being particularly at risk for musculoskeletal symptoms compared to employees working in other occupations. The present study surveyed employees at a call center in the Southeastern United States. Missing data was relatively common in the dataset, and after dropping respondents who had missing data on all values of interest (N=40), we were left with a sample of 352 respondents. Descriptive statistics of the sample can be found in Table 1.

Table 1: Baseline Covariates

Covariate	Notes	Hypothesis
Shift	Time when subject starts to work everyday	subject who has early schedule is more likely to gain weight
Age	Age of subject	the older the subject is, the more likely the subject gain weight(since older people tend to not do exercise)
height	Height of subject	height may not have huge impact on weight gain
pounds_gained	How much weight has the subject gained since starting the job (potential response?)	
BMI	Body mass index (potential response)	
VigexTime	Vigorous MET-minutes used per week	the longer time subject spends on doing vigorous activities per week, the less likely the subject gain weight

ModexTime	Moderate MET-minutes used per week	the longer time subject spends on doing moderate activities per week, the less likely the subject gain weight
WalkexTime	Walk MET-minutes used per week	the amount of time subject spends to walk may not have huge impact on weight gain
Total_Met_Time	total of vigorous, moderate and walking MET-minutes used per week	the longer time subject spends on doing exercises per week, the less likely the subject gain weight
weightgain	if the subject had any weight gain since starting job(yes/no)	
gender	gender of the subject (Male/Female)	gender of the subject may have very little effects on weight gain

Specific Aims

Specific Aim #1: Summarize the descriptive differences between weight gain and demographic and employment measures

1.1 Examine the distributions of two potential dependent variables (a binary measure of weight gain over the survey period and a linear measure of pounds gained over the study period).

1.2 Compare the population of workers who gained weight compared to workers who did not gain weight across social and employment characteristics of interest. For categorical variables (shift time and gender), we will consider Pearson's χ^2 tests for significance. For continuous variables (total MET exercise time and age), we will use t-tests for significant differences.

1.3 Produce histograms for age and MET exercise time to examine the distributions of the continuous predictors. We will also produce histograms comparing group differences by weight gain.

1.4 Investigate missingness in the dataset through regression models predicting missingness on at least one of the predictor variables.

Specific Aim #2: Identify individual level factors associated with weight gain during the study period. We will use two regression methods capable of adjusting for missing data .

2.1 Using the dependent variable selected in SA2, we will use full information maximum likelihood estimation to examine the direction and magnitude of the association between the independent variables and a measure of weight gain. We may consider transformations of the variables if needed. We will also use random forest to analyze the data, and see if the conclusions from random forest match the conclusions from the FIML model.

Conclusion

In general, we want to find what variables affect the chance of gaining weight and how these variables affect the weight gain. We will select appropriate variables as responses first, then produce plots to illustrate the relationships between response and dependent variables that are listed in the Baseline Covariate Table. We will use the knowledge obtained from graphs to select appropriate predictors, and apply full information maximum likelihood regression to examine the relationship between responses and predictors. We will also apply random forest to analyze the same data set, and see if these two models give us consistent results.

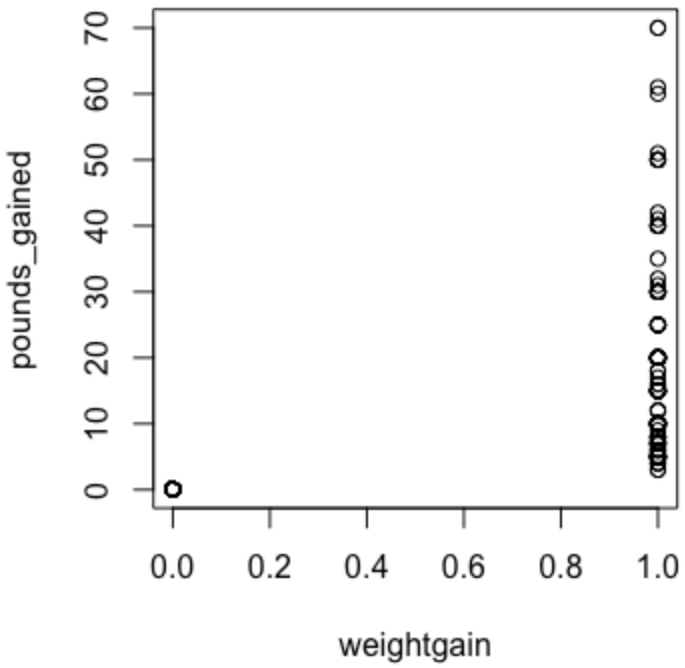
Appendix Table: descriptive statistics

	Statistic	**Analysis Pop**	**Missing**	**All (n=)**
Age	Mean = 33.76	322	30	352
	Sd = 9.89			
	Median = 31			
Gender		347	5	
Female	248 (71%) *			
Male	99 (28%)			
Height (inch)	Mean = 66.64	331	21	

	Sd = 4.04			
	Median = 66			
bweight (lb)	Mean = 178	268	84	
	Sd = 44			
	Median = 170			
BMI	Mean = 27.82	253	99	
	Sd = 6.1			
	Median = 27			
WalkexTime (min/week)	Mean = 123	352	0	
	Sd = 212			
	Median = 60			
Modextime (min/week)	Mean = 74	351	1	
	Sd = 140			
	Median = 30			
VigexTime (min/week)	Mean = 76	352	0	
	Sd = 114			
	Median = 27			
Total_Met_Min (min/week)	Mean = 1306	351	1	
	Sd = 1553			

	Median = 822			
weightgain	No = 111 (32%) *	348	4	
	Yes = 237 (68%)			
pounds_gained	Mean = 11	342	10	
	Sd = 13			
	Median = 8			
shift		348	4	
8am	115(32.7%)			
9am	56(15.9%)			
10am	50(14.2%)			
11am	44(12.5%)			
12pm	14(4%)			
1pm	8(2.3%)			
2pm	15(4.3%)			
Other	15(4.3%)			

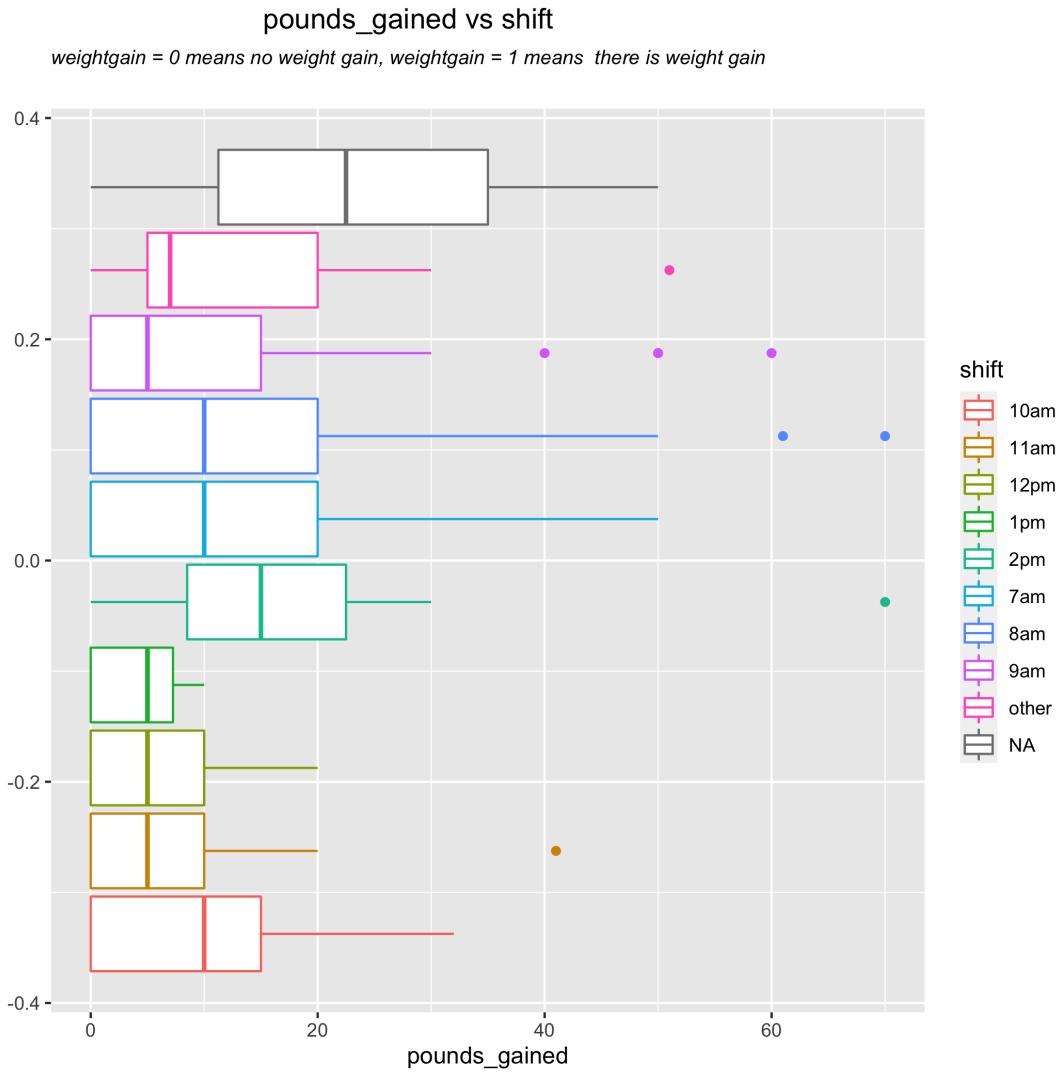
Plot: pounds_gain vs weightgain



The plot above illustrates the relation between variable pounds_gained(lbs) and weightgain. Pounds_gained is a continuous variable. It represents how much weight the subjects have gained. weightgain is binary variable. When weightgain = 0 it means no weight gain. When weightgain = 1 it means there is weight gain.

We can see that these two variables give us consistent information. when weight gain = 0, we have no weightgain thus there is not pound_gained. when weightgain = 1, we have weight gain, thus the pound_gained is not 0. We should not include these two variables in the same model since they are highly correlated.

Plot2: pounds_gained vs shift



From the boxplot above, we can see that factor variable shift has effects on variable pounds_gained. Notice that subjects who have missing value in shift tend to gain more weight than the subjects who report their shift. This indicates that the value in shift may not miss randomly.